

博士論文審査報告書

論文題目

Profiling Technologies for P2P
Data Sharing Communities

P2P型データ共有コミュニティ実現へ
向けたプロファイリング技術に関する研究

申請者

井口	誠
Makoto	Iguchi

情報・ネットワーク専攻

情報システム工学研究

2009年2月

ネットワークを活用したデータの共有には、これまでクライアント・サーバ型のモデルが多用されてきたが、最近になって P2P 型のデータ共有モデルが普及している。P2P 技術を使うことにより、人間社会で情報共有する方法が格段に進歩した。この技術を用いれば、情報を共有する人々がコミュニティを構成して、多種多様なデータに自由にアクセスすることができる。今後とも P2P 型のデータ共有モデルが広く使われるものと予想される。その一方で、P2P 型のデータ共有が大規模にオープンに行われるようになると、幾つかの課題が明らかになってきた。それはユーザの信頼性に関わる問題である。社会の利益に反するような悪意のあるユーザが出現して、虚偽のデータを流布したり、ウィルスやトロイの木馬のような害のあるソフトウェアを忍び込ませるかもしれない。また悪意のあるユーザが他のユーザの個人情報情報を容易に入手する可能性が無視できなくなる。

本論文は、P2P 型のデータ共有モデルの信頼性を確保する技術をプロファイリングに基づいて提案している。すなわち評価の高いユーザを特定する技術、共有すべき評価の高いデータを特定する技術、ネットワークからの不正侵入を早期に検知する技術、個人の秘密を保護した状態で安全に情報共有を実現する技術を、いずれもプロファイリングを用いて実現している。本論文の成果を用いることにより、P2P 型のデータ共有の信頼性を高めることができる。

第 1 章は研究の背景としてクライアント・サーバ型のモデルから P2P 型への発展を概観した上で、本研究の意義を述べている。

第 2 章では P2P 技術でデータを共有するコミュニティにおいて、高い評価を得るユーザを特定するためのプロファイリングを行う方法を提案する。もしコミュニティにおいて高い評価を得ているユーザを容易に特定できるのであれば、そのようなユーザとデータを交換することが推奨される。しかし、膨大な人数のユーザの中から、高い評価を得ているユーザを特定することは容易ではない。本研究はユーザの過去の行動を表現するプロファイリングを用いて、そのユーザの評価を行う。ここではユーザが 2 つの役割を果たす。1 つ目はユーザがデータを提供する役割である。2 つ目はユーザがデータを評価する役割である。多くのユーザが高い評価を与えたデータが存在するときには、そのデータを提供したユーザを高く評価する。これはデータ提供者としてのユーザのプロファイルで表現される。また評価の高いデータを正當に高いと評価したユーザを評価者として高く評価する。これはユーザの評価者としてのプロファイルで表現される。

ユーザのプロファイルと同時に、共有されるデータにも評価を付ける。ユーザのプロファイルとデータの評価とが相互に影響を及ぼす。例えば、提供者として高い評価のユーザが提供したデータの評価は高くなる。また評価者として高い評価のユーザが positive に評価したデータの評価は高くなる。本研究で示した方法は、Napster や Bittorrent のような集中管理型の P2P で

も、Gnutella や Freenet のような分散型の P2P でも適用することができる。また提案した方法が頑健(robust)であることが必要である。これを次のように実現する。悪意のあるユーザが、他のユーザの評価や共有データの評価を意図的に操作するときには、自らの評価が高くなければ影響を及ぼすことができない。悪意のあるユーザが自らの低い評価を捨てて、新規のユーザとして評価の初期値を得ようすることが考えられるので、新規のユーザの評価の初期値を小さくしておく。

第3章はネットワークを経由して不正侵入する行為を検知するためのプロファイリング技術を提案している。P2P 型のデータ共有において、もし悪意のあるユーザとデータ交換をすると、トロイの木馬やウィルスを含むようなデータを送り込まれるかもしれない。そのような事態に陥った場合に、システムが不正に侵入されたことを早期に検知することができれば、被害を最小限に食い止めることができる。ただし不正侵入を早期に検知することは、この分野の大きな課題である。本研究で提案する方法は、TCP/IP のネットワークの通信状態を解析して侵入を検知する。具体的にはパケットを収集して分析する。ここで測定するのは、通信の継続時間、上りと下りの両方向のパケット数、両方向のデータのバイト数、パケット当たりのバイト数である。TCP/IP のアプリケーションプロトコルごとに割り当てられているポート番号に注目して、2 種類のプロファイルを作成する。プロファイルの1つ目は短期のポート毎の観測に基づく。他方は長期のポート毎の観測に基づくプロファイルである。短期のプロファイルが現在のネットワークの活動を表現する。長期のプロファイルは過去のネットワークの活動をまとめたものに相当する。この2つのプロファイルを照合して、両者が一致すれば現在の状態は正常であると判断できる。もし両者に相違があれば現在の状態を異常と判定する。異常時にはユーザに警告を発する必要がある。本研究ではプロファイルを作成する具体的な方法を提案し、プロファイルを照合して差異を判定する方法を提案している。

本研究で提案した方法の評価するために、fingerd のトロイの木馬、FTP bounce attack などの既知の不正侵入で試してみると、長期のプロファイルが安定に作成できること、さらに長期のプロファイルと短期のプロファイルを照合することで、これらの不正侵入の検知が出来ることが実証された。

第4章はユーザの匿名性を保った状態で、P2P 型のデータ交換を可能にする方法を提案している。ここで用いるプロファイルは各ユーザの嗜好を反映したものである。例えば Web ページの閲覧履歴をプロファイルとして表現する。この種の情報はユーザのプライバシーに属するものと考えるべきである。本研究では、類似のプロファイルを持つユーザが、匿名性を保った状態でデータ交換できる方法を提案する。提案する方法の特徴は、ユーザの情報を忠実に伝搬するのではなく、途中のユーザが自分のデータとすり替えてしまうことである。つまり嘘(bogus)の情報を意図的に混入することである。そのよ

うにしてもデータ交換の能率を低下させることはない。

具体的な実現法は次のようになる。ユーザをグラフのノードとする。そのユーザのプロファイルに類似しているプロファイルを持つユーザを捜して、2つのユーザ（ノード）をリンクで接続する。この操作を反復して、類似のプロファイルを持つノードを接続していく。この時に最初のノードから2番目のノードに接続するときには正確にプロファイルを伝えるが、2番目のノードから3番目のノードに接続するときには、2番目のノードは自分のプロファイルではなく、最初のノードのプロファイルを2番目のノードのプロファイルであるかのように伝える。この方法を採用すると、どれが最初のノードであるか分からなくなる。

ここで提案した方法を評価するために、幾つかのシミュレーションを行った。その結果によれば、ここに提案する方法が匿名性を保証しつつ、能率の良いデータ交換を実現していることが分かる。なおユーザをリンクで接続する際には、無限にリンクを延長するのではなく、TTL (Time To Live) というカウンタを設ける。リンクが1段延びると TTL のカウントを1だけ減らす。シミュレーションにより TTL の効果を測定してみると、ユーザの匿名性は TTL の影響をあまり受けない。ただし TTL の値が大きい方が、多くのユーザの間でデータ共有が行われることが分かる。

第5章は結論を述べている。本研究で用いた3つのプロファイリング技術を用いると、P2P型のデータ共有コミュニティにおいて評価の高い信頼されるユーザを特定することができる。また評価の高いデータが分かる。ネットワークの活動を表現するプロファイルにより、ネットワークからの不正侵入を早期に検知することができる。さらに匿名性を保持したままでプロファイルの類似しているユーザとの間で能率良くデータ交換をすることができる。

以上のように本論文の研究成果はプロファイリングの技術を用いて信頼性のあるP2P型のデータ共有を実現しており、社会の基盤となる情報ネットワークの安全な構築をする上で極めて有用である。よって本論文は博士(工学)早稲田大学の学位論文として価値あるものと認める。

2009年1月

審査員	主査	早稲田大学教授	工学博士（東京大学）	後藤 滋樹
		早稲田大学教授	博士(工学)早稲田大学	菅原 俊治
		早稲田大学教授	博士(工学)早稲田大学	山名 早人